BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

EXHIBIT 12

G. SCHNEIDER ET AL

BIOCHEM. BIOPHYS. RES. COMMUN. 194 951-959, 1993

EXHIBIT 12

ANALYSIS OF CLEAVAGE-SITE PATTERNS IN PROTEIN PRECURSOR SEQUENCES WITH A PERCEPTRON-TYPE NEURAL NETWORK

Gisbert Schneider, Sylvia Röhlk, and Paul Wrede+

Freie Universität Berlin, FB Physik, AG Biophysik, Arnimallee 14, 1000 Berlin 33, Germany

Received June 1, 1993

SUMMARY: A method for feature extraction from protein sequences has been developed which is based on an artificial neural filter system. Amino acid sequences are analyzed with regard to physicochemical residue properties. This alternative representation of a sequence allows an interpretation of the networks' weight values in a comprehensive and biochemically meaningful way by displaying the optimized network weights in Hinton diagrams. Signal peptidase cleavage sites of *E.coli* periplasmic proteins, human mitochondrial precursors and chloroplast precursors from spinach have been investigated. The network for *E.coli* periplasmic protein precursors classified both training and test data with 100% accuracy. The interpretation of its network weights clearly confirms the "-3,-1 rule" and the existence of a hydrophobic core region starting at position -6. Further striking features and dominant positions can be found for all three types of cleavage sites.

Many nuclear encoded proteins destined for organelles like mitochondria and chloroplasts are synthesized as precursor proteins in the cytoplasm of the cell. Proteins targeted to the organelles are processed inside of them by special proteases which cleave off N-terminal targeting sequences (1-6). The cleavage-site region of precursor proteins consists of amino acid sequences showing no or only little homology (3, 7, 8). Nevertheless, statistical investigations revealed that the cleavage-site region is locally encoded and thus a good candidate for the development of prediction systems. A clearly defined cleavage-site region exists for *E.coli* precursor proteins (7, 9) as well as for secretory proteins translocated through the endoplasmic reticulum membrane of eukaryotes and for many organelle types (10-12). Since many new sequences will be compiled in the future by sequencing of whole chromosomes from yeast (13) and human, among other functional important sequences the classification of targeting peptides and their cleavage-sites of nuclear encoded organelle proteins will be helpful for understanding genome organization. Until now, alignment procedures and statistical approaches led to the discovery of several rules describing cleavage-site regions already, e.g. the well known "-3, -1 rule" (7, 9) for eubacterial and eukaryotic secretory targeting signals. Using these rules for prediction leads to accuracies less than 100% which still is not sufficient for an accurate analysis of chromosome organization. Reliable prediction systems for the detection of cleavage-sites exist only for secretory signals of eubacterial and eukaryotic precursors (1, 14, 15, 16). It has been reported that artificial neural networks are - in principle - able to recognize complicated sequence patterns automatically, i.e. without giving biased instructions by the investigator (15-18). For an identification of locally encoded sequence patterns improved results for sequence classification and prediction can already be

+To whom correspondence should be addressed.

0006-291X/93 $4.00

obtained with a Perceptron-type neural network although it still has limitations due to its architecture (19). In this work we applied a Perceptron-type neural network to the identification of cleavage-site features which allows a property-based analysis of protein sequence data combined with a neural network feature extraction (Figure 1). Eubacterial signal peptides (targeting signals for export and secretion), chloroplast transit peptides, and mitochondrial transit peptides were analyzed. In contrast to most earlier approaches to this problem, the cleavage-site sequences are described in terms of four physico-chemical amino acid properties which have been identified to be useful for cleavage-site analysis (15, 16, 20): Hydrophobicity (21), hydrophilicity (22), polarity (23), and side-chain volume (24) (Table 1). A complete description of amino acid sequences in terms of physico-chemical properties will be an ultimate necessity to understand the function and structure of proteins, but unfortunately such a description is not possible today. Thus, we focussed on only four amino acid properties which appear to be important. This data representation is thought to reveal characteristic cleavage-site features which cannot be found looking at the sequence letter code.

An evolutionary computing algorithm, the simple Evolution Strategy (25), was used for network training instead of the commonly used generalizing delta rule (26). Evolutionary algorithms are efficient optimization techniques which have been shown to be very useful and produce reliable results when applied to artificial neural networks (15, 16). A dominant feature of a neural network system is its ability to determine which positions and which residues are important for a certain protein structure or function. Most common for the analysis of sequence data is a three-layer feedforward network architecture which could be shown to be able to approximate any continuous input-output relation (27, 28), e.g. the identification of cleavage-sites in precursor sequences. Although these multilayer networks appear to be well suited for the development of prediction systems they lack the possibility to explicitely show what the important sequence features are. On the one hand this is due to the non-symbolic technique itself, on the other hand many investigations using neural networks are based on a sequence representation in terms of binary numbers representing characters rather than physico-chemical property values. To overcome this disadvantage we used a two-layer network architecture for the analysis of cleavage-site features which allows a property-based analysis of protein sequence data combined with a neural network feature extraction. The obtained optimized networks have been analyzed using Hinton-diagrams which offers the possibilty to interprete a network's weight values w_{ij} (Figure 1).

METHODS

Data selection and preparation: 24 *E.coli* periplasmic protein precursors, 27 chloroplast precursor sequences from spinach, and 39 human mitochondrial precursor proteins with experimentally confirmed cleavage-sites were selected from the SwissProt database, Rel. 20. The sequences were randomly divided into training sets and test sets following a ratio of 7:3 for every type of precursor. This resulted in 17 *E.coli* 19 chloroplast, and 27 mitochondrial training sequences. These data were used for training of the Perceptron system. The remaining test set sequences were used for an evaluation of the networks generalization ability by measuring the prediction accuracy of the optimized networks when applied to the corresponding test set. We are aware that e.g. cross-validation tests are more precise compared to an evaluation with regard to a single test set. Our aim was not to establish a new prediction method. Rather, we want to present a useful technique for protein data analysis using physico-chemical amino acid properties. A complete list of the data sets including names and sequences of the precursors is available from the authors on request. For training of the networks on the detection of cleavage-site features the training data were restricted to sequence strings covering 10 residues of the targeting sequence and 2

residues of the N-terminal end of the mature protein (positions -10 to +2). These 12-residue windows served as positive examples for network training. For every positive example 3 negative examples were selected randomly from the whole precursor sequence, e.g. the whole training data for the *E.coli* cleavage-site consisted of (17 positive + 51 negative = 68) sequence windows.

Network architecture and training technique: A model of our Perceptron-type neural network is given in Figure 1: The amino acids of a given sequence window are numerically described by four physico-chemical side chain properties: Hydrophobicity (21), hydrophilicity (22), polarity (23), and volume (24), leading to a (12 x 4) property matrix. The scales were normalized to give comparable values between 0.0 and 1.0 (Table 1). Every value of the input property matrix x_{ij} is connected by a weight factor w_{ij}. The single output unit calculates the network's output value ($0.0 < y < 1.0$) for a given sequence window using a sigmoid function $F(x) = 1 / (1 + \exp(-x))$ as transfer function:

$$y = F(\sum_{i,j} x_{ij} w_{ij}).$$

In contrast to the classical Perceptron system (26), the output unit of our network employs a sigmoidal transfer function. Therefore, the separation of positive and negative examples of a training set is not restricted to linear separation (19). The task was to find the correct weight values w_{ij}, which allow an optimal separation between positive and negative examples. For this, a (1,100)-evolution strategy with adaptive stepsize control was used (25). This optimization method employs a systematic top down search in the feature space imitating the natural process of repeated mutation and selection. Here, the weight values and the stepsize in a learning cycle were the parameters for this mutation-selection procedure (generate-and-test cycle) starting with random values. This evolutionary algorithm has already been successfully applied to the optimization of neural networks for pattern recognition in protein sequences (15,16). It is described there. The best weight values of a learning cycle are selected following an external quality function. We used the minimization of the square error Δ^2 for supervision of the learning process:

$$\Delta^2 = \sum_{p=1}^{n} (t_p - y_p)^2.$$

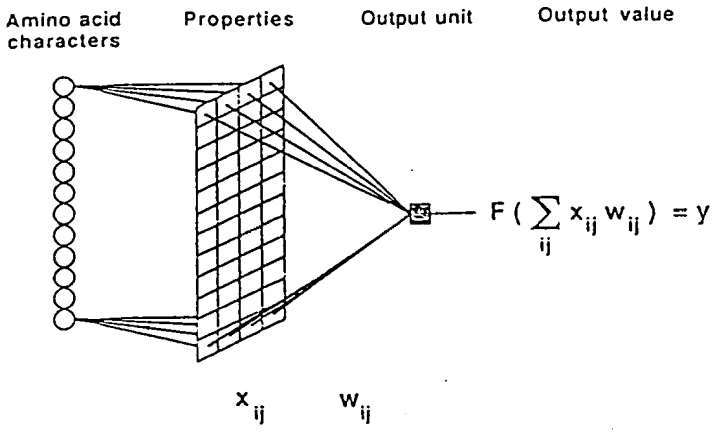

Figure 1. Model of the Perceptron architecture. Amino acid sequences are represented in terms of four physico-chemical properties x_{ij}. The single output unit uses a sigmoidal function F(x) as transfer function. The output value y for a sequence window of 12 residues is calculated by the given formula. For clarity, only the connections for the first and the last residue of the input window are drawn. In total, 48 network connection weights w_{ij} had to be optimized.

Table 1. The normalized property scales used for the sequence description. References, see text.

Amino acid	Hydrophobicity	Hydrophilicity	Polarity	Volume
Ala	0.8688	0.4531	0.0	0.1699
Arg	0.0	1.0	1.0	0.6756
Asn	0.4688	0.5625	0.6500	0.3435
Asp	0.1938	1.0	0.9558	0.3041
Cys	0.8938	0.3750	0.2846	0.2886
Gln	0.5125	0.5625	0.6788	0.4997
Glu	0.2563	1.0	0.9596	0.4669
Gly	0.8313	0.6875	0.0	0.0
His	0.5813	0.4531	0.9923	0.5551
Ile	0.9625	0.2500	0.2500	0.6357
Leu	0.9438	0.2500	0.2500	0.6357
Lys	0.2188	1.0	0.9519	0.6470
Met	0.9813	0.3281	0.2750	0.6130
Phe	1.0	0.1406	0.6731	0.7740
Pro	0.7563	0.5313	0.3038	0.3733
Ser	0.8063	0.5781	0.3212	0.1723
Thr	0.8438	0.7188	0.3192	0.3339
Trp	0.8875	0.0	0.4038	1.0
Tyr	0.7250	0.1719	0.3096	0.7961
Val	0.9313	0.2969	0.2500	0.4764

Here, n is the number of sequence windows in the training set, t_p is the desired output value (1.0 for a positive example, 0.0 for a negative example), and y_p is the actual network output for the training sequence p. Starting with random weight values this quality function served as a heuristic for a systematic search fo optimal values. To reduce the effect of overlearning, i.e. a specialization of the network on the training se sequences rather than the extraction of a general feature, a second simple statistical quality function Q (29) was used to determine the termination time of the learning process. Q calculates the prediction accuracy of the network:

$$Q = \frac{P + N}{tot}.$$

Here, P is the positive correct prediction, N is the negative correct prediction, and tot is the total number of investigated residues. If Q reaches the value 1.0 (no prediction error) the learning process will terminate. A maximum of 200 learning cycles (generations) was allowed.A neural filter network for reliable recognition of cleavage-site features must show a high Q value (max. 1.0) and a low absolute square error Δ^2 of i output value. Overprediction of cleavage-sites (false positive prediction) will occur if the network ha "learned" a too general feature, underprediction (false negative prediction) will occur if a special training set feature is used for prediction. In both cases the neural filter system does not show a good generalization ability.

RESULTS

Network training: For the *E. coli* targeting peptide cleavage-site a feature could be found leading to 100% correct classification of training and test set examples (Q) with a final output error of 0.74 (Table 2 The features found for chloroplast and mitochondrial precursor cleavage-sites allowed 99% and 97% correct classification of the training data, respectively. The Q values for the independent test sets are lower (72% and 79%), the corresponding ouput errors are higher than for the *E. coli* sequences (1,17 and 4.89) indicating that no generalizing feature could be found.

Table 2. The filters for the three investigated types of cleavage-sites were evaluated with regard to their learning success (prediction accuracy). Q measures positive and negative correct prediction in the training set (Q_{train}) an an independent test set (Q_{test}). Δ^2 is the final square error of the network output value. Training and test sets for learning consisted of 12-residue windows.

Data	$Q_{training}$	Q_{test}	Δ^2
E.coli	100 %	100 %	0.74
Chloroplasts	99 %	72 %	1.17
Mitochondria	97 %	79 %	4.89

A critical step in the computer experiments was the choice of the size of the input layer. The number of 12 units was chosen because it appears to an ideal window size for eubacterial cleavage-sites (1). In fact, a prediction accuracy (Q) of 100% could be achieved for the *E.coli* sequences (Table 2). Thus, we conclude that the signal for cleavage by periplasmic signal peptidase is locally encoded and completely described by a short stretch of amino acids. Here, the sequence representation in terms of the four physico-chemical properties (Table 1) was sufficient for the extraction of a generalizing cleavage-site feature. Chloroplast signals appear to be more locally encoded (Q=99%) than mitochondrial cleavage-sites (Q=97%), although this statement is not very well supported. It is striking that in these two cases the data preparation was not ideal. Several empirical parameters of the neural network system, e.g. the transfer function or the size of the input matrix, should be optimized for further analysis of the organelle sequences, too. Recently, a systematic approach for an optimization of neural network parameters which is also based on evolutionary computing algorithms has been proposed (30).

Interpretation of the Hinton-diagrams: An important aspect of our Perceptron-type network architecture is the use of an amino acid property matrix as input instead of the common 20-bit binary number coding for a residue. The sequence representation as a property matrix of real numbers allows a quasi-symbolic interpretation of the network weights in a comprehensible and biochemically meaningful way. A simple way to do this is a graphical representation similar to a Hinton diagram (Figure 2). The obtained weight values were normalized and splitted into four groups, see legend Figure 2. We interprete extreme weight values - indicated by black or white squares - to show the most important sequence positions. Nevertheless, a whole diagram must be regarded as the respective cleavage-site feature for *E.coli* (Figure 2A), chloroplasts (Figure 2B) and mitochondria (Figure 2C). It must be stressed that only the properties hydrophobicity and volume are not correlated, while all other properties are substantially correlated. For this reason, special care must be taken when interpreting the Hinton diagrams.

Description of *E.coli* signal peptide cleavage-site patterns: *E.coli* signal peptides contain a hydrophobic part beginning at position -6 (Figure 2A). Here, small residues appear to be important, too. Additional preference for hydrophobic residues can be found at the position -1 and -3, where position -3 appears to be less hydrophobic than -1. The positions -3 and -1 are also dominant with regard to the properties polarity and volume, small apolar amino acids seem to play an essential role for the cleavage-site

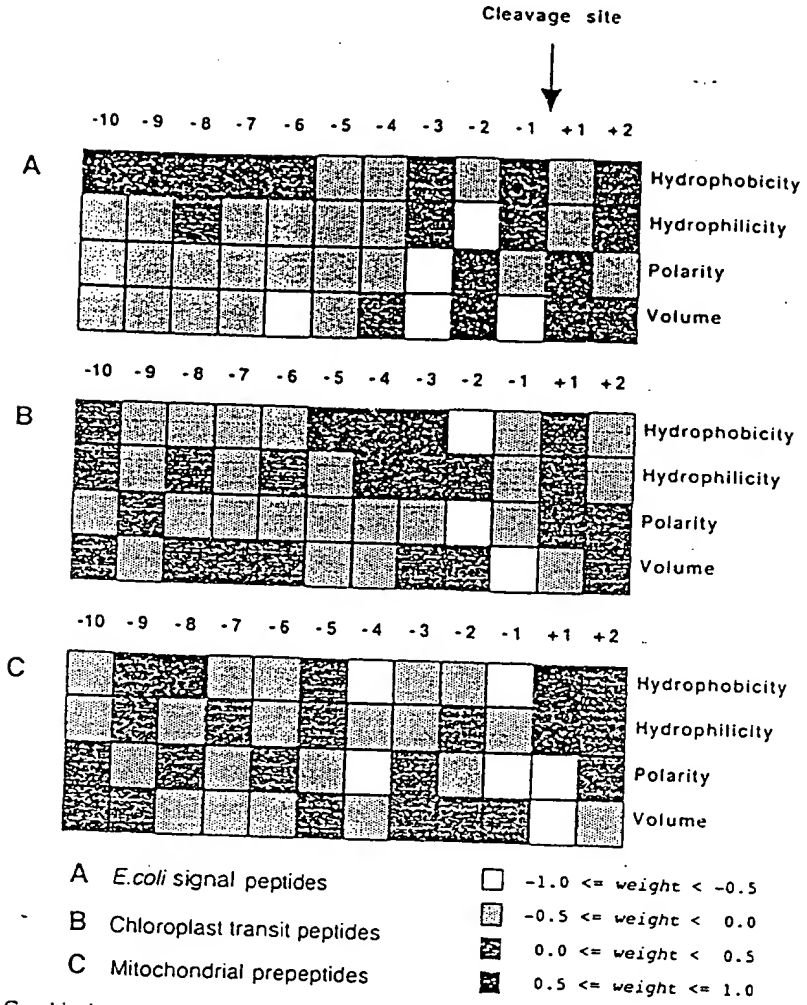

Figure 2. Graphical representation of the networks' weight values w_{ij} (Hinton-diagram). The (12 x 4) matrices are thought to represent the cleavage-site features extracted by the neural filters for the three types of precursors (A, B, C). The cleavage-sites are indicated by an arrow. The mature protein starts at position +1 (N-terminal end of the mature protein).

signal. In position -2 there is a requirement for large non-hydrophilic side chains. We regard position -2 as being "complementary" to -3 and -1. In general, there are two predominant sections around the signal peptidase cleavage-site: A mainly hydrophobic part starting at -6 and the positions -3 to -1. Previous investigations already identified this "hydrophobic core" and the "-3,-1 rule" (7, 9, 15, 20, 31). By help of our weight interpretation a more precise characterization of the quantitative influence of single positions an amino acid properties is possible now. Polar residues at the N-terminus of the mature protein (+1) appear to contribute to the cleavage-site signal, too.

Description of chloroplast transit peptide cleavage-site patterns: The cleavage-site regions in chloroplast transit peptides (Figure 2B) do not have two separate regions as in E.coli precursors. Besides a

cluster of extreme values at the positions -5 to -1 no important parts can be identified. Hydrophobic residues appear to be dominant at -3 and -5, separated by a strongly hydrophilic side chain at -4. At position -2 a lack of hydrophobicity and polarity, at -1 small residues seem to be crucial.

Description of mitochondrial prepeptide cleavage-site patterns: The positions -8, -4, -1 and +1 show extreme weights in the mitochondrial precursor sequences (Figure 2C). Position -8 appears to be mainly hydrophobic. The positions -4 and -1 clearly lack hydrophobic and polar amino acids. At position +1 small apolar side chains appear to be essential. The observation of a positive charge (Arg) at position -2 (11) cannot be confirmed.

DISCUSSION

Cleavage-site features for the leaderpeptidase of *E.coli* periplasmic protein precursors could be detected by the application of a simple artificial neural network, a Perceptron. Different features of cleavage-site regions from the nuclear encoded precursor proteins destined for the inner mitochondrial membrane and those of the chloroplast transit peptides were recognized by an accuracy of 97% resp. 99%, while for the *E.coli* 100% accuracy was obtained with the training data. Systematic investigations described earlier led to the conclusion that a description of amino acid sequences by the four physico-chemical properties hydrophobicity, hydrophilicity, polarity and side-chain volume was appropiate in describing the crucial features of the cleavage-site regions which were recognized by the Perceptron system (15, 16, 20). The appropiate choice of physico-chemical properties for describing amino acid sequences depends on the particular function of the sequence. The cleavage-site function can be described by the four selected properties to be accurately recognized by our neural network.

Features of *E.coli* signal peptide cleavage-sites: The Perceptrons' weight factors are interpreted to correlate with the influence of the property in the particular sequence position. This assumption is only valid if the prediction in the training- and test-sets are 100%. The learned weight values can be graphically displayed by a Hinton-diagram (Figure 2). The interpretations of the Hinton-diagrams should agree with already known cleavage-site rules. *E.coli* signal peptides for protein export and secretion have been analyzed very intensively and some reliable rules for cleavage-site prediction exist (7, 9). The results obtained with the Perceptron system confirm the validity of the "-3, -1 rule" found earlier (3, 7) for *E.coli* periplasmic proteins. This rule implies to small and very hydrophobic amino acids in these positions. For position -2 several amino acids have been suggested, while an analysis of the Hinton-diagram indicates that large amino acids are essential (Figure 2A). Our findings are confirmed by recent cassette-mutagenesis experiments replacing the native cleavage-site of the *E.coli* alkaline phosphatase by a series of idealized nonnatural model sequences (31). Further, position -8 seems to be necessary to slightly disturb the hydrophobic core. Position -5 is exceptional in a sense that no extreme weight values for any of the four properties came up. At -6, proline was suggested to be important inducing an interruption of a possible helix (1). Experimental data disproved the role of a proline (31) as well as biocomputing results by our PROSA-Design algorithm (20). Following the Hinton-diagram, the hydrophobic core starts at position -6, where a small and somewhat hydrophobic amino acid is essential (e.g. Ala, Gly, Cys, Pro). Since both training and test data are classified with 100% accuracy (Table 2) the obtained features can be regarded as being generally valid.

Features of chloroplast transit peptide cleavage-sites: Chloroplast targeting sequences contain a great number of hydroxylated serine and threonine residues (32). The cleavage-site can be described by a rather loosely defined consensus sequence: (Val/Ile)-X-(Ala/Cys)↓Ala (X means any amino acid), arginines are frequently found in the positions -6 to -10 (33). A comparison with the Hinton-diagram (Figure 2B) confirms the cleavage-site rule (high hydrophobicity at position -3) but in addition position -2 can now be precisely characterized by a non-hydrophobic and non-polar residue which does not suit to arginine (Table 1). The assumed equivalent role of valine and isoleucine at -3 is not surprising because these two amino acids resemble closely in their physico-chemical properties (Table 1). On the other hand, an equivalent role of alanine and cystein at position -1 cannot be explained by a similarity in their properties except for volume. It may be that cystein and alanine have some other properties in common which are not covered by Table 1. At -1 the weight factors of the trained Perceptron did not reach extreme values in three properties except for volume (Figure 2b). A very small size will fit for alanine, but less well for cysteine which is larger than alanine, glycine and serine. According to the Hinton-diagram, a new chloroplast cleavage-site feature can be identified which could be described by a "-3,-5 box" that allows very hydrophobic amino acids only, while in position -4 very hydrophilic amino acids seem to be preferred. The chloroplast cleavage-site rule based on the known experimental and statistical results - including those described here - does not seem to be homologous to the rule found for eubacterial cleavage-sites. The Perceptron did not find a generalizing feature as the quality index for the test set is only 72% (Table 2). Since all hitherto carried out analyses of chloroplast targeting sequences use only a single set of sequences - which is analogue to our single training set - and no test set at all, it is not surprising that our results do not contradict known facts. Further optimization of the neural network architecture is essential to obtain generally valid features (16, 30). Anyway, it was possible to obtain several results consistent with known experimental facts and apart from that some more possible characteristics have been found. Site directed mutagenesis experiments should clarify these uncertainties.

Features of mitochondrial prepeptide cleavage-sites: Mitochondrial cleavage-sites for the recognition by the major matrix protease (protease I in higher eukaryotes) have been analyzed with a greater success by the Perceptron than the chloroplasts' transit peptide cleavage-sites (Table 2). The obtained square error of 4.89 is rather high compared to the *E.coli* result. Thus, some care has to be taken with the interpretation of the Hinton-diagram (Figure 2C). Position -8 is found to be occupied by strong hydrophobic residues which is confirmed by literature (34). Further, in +1 small and apolar amino acids which are also very hydrophobic as well as hydrophilic seem to be necessary. Such a property description fits best to the amino acid glycine according to Table 1. The Hinton-diagram may be interpreted further in the description of a cleavage-site rule by claiming an "-4,-1 box" which is characterized by apolar and non-hydrophobic residues. Several features of the different cleavage-site classes have been identified by our Perceptron system. Especially the choice of a description of primary structures in terms of physico-chemical property scales was a necessary prerequisite and led to interpretable results. It is clearly shown that even a single-unit network is able to extract biochemically comprehensive rules from molecular sequence data. Probably, the results for chloroplasts and mitochondrial cleavage-sites will be improved if more properties are concerned and feature extraction is performed by more powerful multilayer networks.

ACKNOWLEDGMENTS: Prof.'s G. Büldt and H. Schweppe are thanked for generous support, helpful discussions, and working facilities. G.S. is supported by the Fonds der Chemischen Industrie (FCI). The project has been granted by the Deutsche Forschungsgemeinschaft (DfG, Sfb 312).

REFERENCES

1. von Heijne, G. (1986) Nuc. Acid Res. 14, 4683-4690.
2. Roise, D. & Schatz, G. (1988) J. Biol. Chem. 263, 4509-4511.
3. Pugsley, A.P. (1989) *Protein Targeting*. Academic Press, San Diego, C.A.
4. Smeekens, S., Weisbeck, P. & Robinson, C. (1990) Trends Biochem. Sci. 15, 73-76.
5. de Boer, A.D., Bakker, H., Lever, A., Bouma, T., Salentijn, E. & Weisbeek, P.J. (1991) EMBO J. 10, 2765-2772.
6. Schatz, G. (1993) Protein Sci. 2, 141-146.
7. von Heijne, G. (1983) Eur. J. Biochem. 133, 17-21.
8. Watson, M.E.E. (1984) Nuc. Acids Res. 12, 4155-4174.
9. Perlman, D. & Halvorson, H.A. (1983) J.Mol. Biol. 167, 391-409.
10. Schatz, G. (1987) Eur. J. Biochem. 165, 1-6.
11. Hendrick, J.P., Hodges, P.E. & Rosenberg, L.E. (1989) Proc. Natl. Acad. Sci. U.S.A. 86, 4056-4060.
12. Hartl, F.U., Pfanner, N., Nicholson, D.W. & Neupert, W. (1989) Biochim. Biophys Acta 988, 1-45.
13. Olivier, S.G. et al. (1992) Nature 457, 38-46.
14. Ladunga, I., Czako, F., Csabai, I. & Geszti, T. (1991) Comput. Appl. Biosci. 7, 485- 487.
15. Schneider, G. & Wrede, P. (1992) Endocyt. Cell Res. 9, 1-12.
16. Schneider, G. & Wrede, P. (1993) J. Mol. Evol. 36, in press.
17. Holley, L.H. & Karplus, M. (1989) Proc. Natl. Acad. Sci. U.S.A. 86, 152-156.
18. Hirst, J.D. & Sternberg, M.J.E. (1992) Biochemistry 31, 7211-7218.
19. Minsky, M.L. & Papert, S.A. (1988) *Perceptrons*. MIT Press, Cambridge, M.A.
20. Schneider, G. & Wrede, P. (1993) Protein Seq. Data Anal. 5, 227-236.
21. Engelman, D.A., Steitz, T.A. & Goldman, A. (1986) Annu. Rev. Biophys. Biophys. Chem. 15, 321-353.
22. Hopp, T.P. & Woods, K.R. (1981) Proc. Natl. Acad. Sci. U.S.A. 78, 3824-3828.
23. Jones D.D. (1975) J. theor. Biol. 50, 167-183.
24. Zamyatnin, A.A. (1972) Prog. Biophys. Mol. Biol. 24, 107-123.
25. Rechenberg, I. (1973) *Evolutionsstrategie*: Frommann-Holzboog. Stuttgart.
26. Rosenblatt, F. (1958) Psychological Review 65, 386-408.
27. Cybenko, G. (1989) Mathematics of Control, Signals, and Systems 2, 303-314.
28. Hornik, K., Stinchcombe, M. & White, H. (1989) Neural Networks 2, 359-366.
29. Schulz, G.E. & Schirmer, R.H. (1973) *Principles of Protein Structure*: Springer-Verlag, Stuttgart.
30. Lohmann, R. (1992) In *Dynamic, Genetic and Chaotic Programming* (Soucek, B. & the IRIS Group, Eds.), pp. 395-411. Wiley & Sons, New York.
31. Laforet, G.A. & Kendall, D.A. (1991) J. Biol. Chem. 266, 1326-1334.
32. Franzén, L.G., Rochaix, J.D. & von Heijne, G. (1990) FEBS lett. 260, 165-168.
33. Gavel, Y. & von Heijne, G. (1990a) FEBS lett. 261, 455-458.
34. Gavel, Y. & von Heijne, G. (1990b) Protein Eng. 4, 33-37.